

Model Adequacy Checks for Discrete Choice Dynamic Models*

IGOR KHEIFETS[†] AND CARLOS VELASCO[‡]

February 15, 2012

Abstract

This paper proposes new parametric model adequacy tests for possibly nonlinear and nonstationary time series models with noncontinuous data distribution, which is often the case in applied work. In particular, we consider the correct specification of parametric conditional distributions in dynamic discrete choice models, not only of some particular conditional characteristics such as moments or symmetry. Knowing the true distribution is important in many circumstances, in particular to apply efficient maximum likelihood methods, obtain consistent estimates of partial effects and appropriate predictions of the probability of future events. We propose a transformation of data which under the true conditional distribution leads to continuous uniform iid series. The uniformity and serial independence of the new series is then examined simultaneously. The transformation can be considered as an extension of the integral transform tool for noncontinuous data. We derive asymptotic properties of such tests taking into account the parameter estimation effect. Since transformed series are iid we do not require any mixing conditions and asymptotic results illustrate the double simultaneous checking nature of our test. The test statistics converges under the null with a parametric rate to the asymptotic distribution, which is case dependent, hence we justify a parametric bootstrap approximation. The test has power against local alternatives and is consistent. The performance of the new tests is compared with classical specification checks for discrete choice models.

Keywords: Goodness of fit, diagnostic test, parametric conditional distribution, discrete choice models, parameter estimation effect, bootstrap.

JEL classification: C12, C22, C52.

*We thank Juan Mora for useful comments. Financial support from the Fundación Ramón Areces and from the Spain Plan Nacional de I+D+I (SEJ2007-62908) is gratefully acknowledged.

[†]New Economic School, Moscow.

[‡]Department of Economics, Universidad Carlos III de Madrid.

1 Introduction

Dynamic choice models are important econometric tools in applied macroeconomics and finance. These are used to describe the monetary policy decisions of central banks (Hamilton and Jordá, 2002; Basu and de Jong, 2007), for recession forecasting (Kauppi and Saikkonen, 2008; Startz, 2008) and to model the behavior of agents in financial markets (Rydberg and Shephard, 2003). In the simplest framework, a binary dynamic model explains the value of an indicator variable in period t , $Y_t \in \{0, 1\}$, in terms of an information set Ω_t available at this period. Then Y_t conditional on Ω_t is distributed as a Bernoulli variable with expectation $p_t = E(Y_t|\Omega_t) = P(Y_t = 1|\Omega_t) = F(\pi_t)$ where $\pi_t = \pi(\Omega_t)$ summarizes the relevant information and F is a cumulative probability distribution function (cdf) monotone increasing. Typical specifications of the link function F are the standard normal cdf, Φ , and the logistic cdf.

We can describe the observed values of Y_t as $Y_t = 1\{Y_t^* > 0\}$ where Y_t^* is given by the latent variable model

$$Y_t^* = \pi_t + \varepsilon_t$$

and $\varepsilon_t \sim F = F_\varepsilon$ are iid observations with zero mean.

In a general specification π_t is a linear combination of a set of exogenous variables X_t observable in t , but not necessarily contemporaneous, plus lags of Y_t and π_t itself,

$$\pi_t = \alpha_0 + \alpha(L)\pi_t + \delta(L)Y_t + X_t'\beta,$$

where $\delta(L) = \delta_1 L + \dots + \delta_q L^q$ and $\alpha(L) = \alpha_1 L + \dots + \alpha_p L^p$. When $q = 0$, $p = 1$ and $F_\varepsilon = \Phi$ this leads to the dynamic probit model of Dueker (1997),

$$\pi_t = \pi_0 + \delta_1 Y_{t-1} + X_t'\beta,$$

and if the roots of $1 - \alpha(L)$ are out of the unit circle, π_t can be represented in terms of infinite lags of Y_t and X_t .

Many nonlinear extensions have been considered in the literature, such as interactions with lags of Y_t , to describe the state of the economy in the past,

$$\pi_t = \pi_0 + \delta_1 Y_{t-1} + X_t'\beta + (Y_{t-1} X_t)'\gamma$$

or with the sign of other variables in X_t , both stressing different reaction functions in several regimes defined in terms of exogenous variables at period t . Other specifications consider heteroskedasticity corrections, so that $\text{Var}(\varepsilon_t) = \sigma^2(\Omega_t)$, for example a two regimes conditional variance, $\text{Var}(\varepsilon_t) = \sigma^2(Y_{t-1})$.

In the general ordered discrete choice model, the dependent variable takes $J + 1$ values in a set \mathcal{J} , and the parametric distribution $P(Y_t = j|\Omega_t)$ can be modeled using the unobserved latent continuous dependent variable Y_t^* . In the typical case where $Y_t = j$ if $\mu_{j-1} \leq Y_t^* \leq \mu_j$ for $j \in \mathcal{J}$, $\mathcal{J} = \{0, 1, 2, \dots, J\}$ and $\varepsilon_t \sim F_\varepsilon$, with $\mu_{-1} = -\infty$ and $\mu_J = \infty$, we have that

$$P(Y_t = j|\Omega_t) = F_\varepsilon(\mu_j - \pi_t) - F_\varepsilon(\mu_{j-1} - \pi_t)$$

with $\alpha_0 = 0$.

Forecasting is one of the main uses of discrete choice models. In that case for the calculation of predictions it might be necessary to resource to recursive methods when $\delta(L) \neq 0$. However in almost all situations parameters are unknown, but conditional maximum likelihood (ML) estimation is straightforward given the binomial or discrete nature of data, with typically well behaved likelihoods and asymptotic normal estimates if the model is properly specified. The existence, representation and probability properties of these models have been studied under general conditions by de Jong and Woutersen (2011), who also report the consistency and asymptotic normality of ML estimates

when the parametric model is correct. However, if not, estimates will be inconsistent and predictions can be severely biased.

This leads to the need of diagnostic and goodness-of-fit techniques, which should account for the main features of these models, discrete nature and dynamic evolution. The first property entails nonlinear modeling and renders invalid many methods specifically tailored for continuous distributions. Though the latent disturbance ε_t is continuous and with a well specified distribution, it is unobservable. Simulation methods could be used to estimate the distribution of such innovations, but we follow an alternative route by "*continuing*" the discrete observations Y_t , so that they have a continuous and strictly increasing conditional distribution in $[-1, J]$ given Ω_t . This distribution inherits the dependence on a set of parameters and on a conditional information set and can serve as a main tool to evaluate the appropriateness of the hypothesized model.

Conditional distribution specification tests are often based on comparing parametric and non-parametric estimation as Andrews' (1997) conditional Kolmogorov test, or on the integral transform (see Bai 2003, Corradi and Swanson 2006). The former approach is developed for different data types, while the latter can be used only for data with continuous distribution. The integral transform does not require strong conditions on the data dependence structure, so it is very useful in testing dynamic models. However, applying the integral transform to noncontinuous data will not bring to uniform on $[0, 1]$ series, and therefore this approach can not be applied directly to dynamic discrete choice models. To guarantee that adequacy tests based on the integral transform enjoy nice asymptotic properties we propose the following procedure: first, make data continuous by adding a continuous random noise and then apply the modified conditional distribution transformation to get uniform iid series.

The first step can be called the *continuous extension of a discrete variable* which has been employed in different situations. For example Ferguson (1967) uses some type of extension for simple hypothesis testing, Denuit and Lambert (2005) and Neslezhova (2006) use it to apply a copulas technique for discrete and discontinuous variables. The second step is the probability integral transform (PIT) of the continued variables, which we will call *randomized PIT*. Resulting uniform iid series can be tested using Bai (2003) or Corradi and Swanson (2006) tests. However, in some cases these tests can not distinguish certain alternatives, so we also propose test based on comparing joint empirical distribution functions with the product of its theoretical uniform marginals by means of Cramer-von Mises or Kolmogorov-Smirnov type statistics, developed by Kheifets (2011) for continuous distributions.

In a general setup, we do not know the true parameters, while the integral transform using estimated parameters does not necessary provide iid uniform random variates. Hence asymptotic properties and critical values of the tests with estimated parameters have to be addressed. The estimation effect changes the asymptotic distribution of the statistics and makes it *data dependent*. Andrews (1997) proves that parametric bootstrap provides correct critical values in this case using linear expansion of the estimation effect, which arises naturally under the ML method. The idea of orthogonal projecting the test statistics against the estimation effect due to Wooldridge (1990) has been used in parametric moment tests, see Bontemps and Meddahi (2006). The continuous version of the projection, often called Khmaladze (1981) transformation, was employed in the tests of Koul and Stute (1999) to specify the conditional mean, and of Bai and Ng (2001), Bai (2003), Delgado and Stute (2008) to specify the conditional distribution. These projection tests are not model invariant since they require to compute conditional mean and variance derivatives, and also projections may cause a loss in power. In this paper we apply a bootstrap approach instead. In the case of ordered choice models an extensive Monte Carlo comparison of specification tests has been done by Mora and Moro (2008) in a static cross section context. They study two types of tests based on moment conditions and on comparison of parametric and nonparametric estimates.

Despite that there is some work on nonstationary discrete data models, cf. Phillips and Park (2000), we stress stationary situations, but some ideas could be extended to a more general set up as

far as the conditional model provides a full specification of the distribution of the dependent discrete variable.

The contributions of this paper are following: 1) a new specification test for dynamic discontinuous models is proposed, 2) we show that the test is invariant to the choice of distribution of the random noise added, 3) parameter estimation effect of the test is studied, 4) under standard conditions we show the asymptotic properties of such tests, and 5) since asymptotic distribution is case dependent, critical values can not be tabulated and we prove that a bootstrap distribution approximation is valid.

The rest of the paper is organized as follows. Section 2 introduces specification test statistics. Asymptotic properties and bootstrap justification provided in Section 3. Monte Carlo experiments are reported in Section 4. Section 5 concludes.

2 Test statistics

In this section we introduce our goodness-of-fit statistics. Suppose that a sequence of observations $(Y_1, X_1), (Y_2, X_2), \dots, (Y_T, X_T)$ is given. Let $\Omega_t = \{X_t, X_{t-1}, \dots; Y_{t-1}, Y_{t-2}, \dots\}$ be the information set at time t (not including Y_t). We consider a family of conditional cdf's $F(y|\Omega_t, \theta)$, parameterized by $\theta \in \Theta$, where $\Theta \subseteq R^L$ is a finite dimensional parameter space. We could allow for nonstationarity by permitting the change in the functional form of the cdf of Y_t using subscript t in F_t . Our null hypothesis of correct specification is

H_0 : The conditional distribution of Y_t conditional on Ω_t is in the parametric family $F(y|\Omega_t, \theta)$ for some $\theta_0 \in \Theta$.

For example, for dynamic ordered discrete choice model the null hypothesis would mean that $\exists \theta_0 \in \Theta, \quad \forall j = 0, \dots, J, \quad P(Y_t = j | \Omega_t) = p_j(\Omega_t, \theta_0)$, i.e. that all conditional probabilities are in a given parametric family.

For further analysis, we assume that the support of the conditional distributions $F(y|\Omega_t, \theta)$ is a finite set of nonnegative integers $\{0, \dots, J\}$ and $F(y|\Omega_t, \theta) = \sum_{j \leq y} P_F(j|\Omega_t, \theta)$, where P_F is the probability function at the discrete points.

The first step is to obtain a *continuous version* of Y . For any random variable $Z \sim F_z$ with support in $[0, 1]$ and F_z continuous (but not necessary strictly increasing) define the *continued by Z* version of Y ,

$$Y^\dagger = Y + Z - 1.$$

Then the distribution of the continued version of Y is

$$F^\dagger(y|\Omega_t) = P(Y^\dagger \leq y|\Omega_t) = F([y]|\Omega_t) + F_z(y - [y]) P([y] + 1|\Omega_t), \quad (1)$$

which is strictly increasing on $[-1, J]$. The typical choice for Z is the uniform in $[0, 1]$, so that

$$F^\dagger(y|\Omega_t) = F([y]|\Omega_t) + (y - [y]) P([y] + 1|\Omega_t). \quad (2)$$

The binary choice case renders $F^\dagger(y|\Omega_t) = (y - [y])(1 - p_t)$ for $y \in [-1, 0)$ and $F^\dagger(y|\Omega_t) = (1 - p_t) + (y - [y])p_t$ for $y \in [0, 1]$. Note, that F^\dagger coincides with F in the domain of F . We state next an "invariance property": for our purpose, it does not matter how to continue Y and what distribution F_z of the noise Z to add. The unit support of Z is needed to get a simple expression for F^\dagger in (1), otherwise the resulting distribution will be a convolution $F^\dagger(y|\Omega_t) = \sum_{j=0}^J F_z(y + 1 - j) P(j|\Omega_t)$. Continuation idea has been used to deal with discrete distributions, for example, to work with copulas with discrete marginals as in Denuit and Lambert (2005).

The following proposition generalizes results about the probability integral transform.

Proposition 1 (a) Under H_0 random variables $U_t = F^\dagger(Y_t^\dagger|\Omega_t, \theta_0)$ are iid uniform; (b) Invariant property of randomized PIT: realizations of U_t are the same for any distribution F_z in (1) both under H_0 and under the alternative.

Part (a) is a property of usual PIT with a continuous distribution F^\dagger . Part (b) that realizations of U_t are the same, means the following. Consider continuations of Y_t by arbitrary $Z \sim F_z$ and uniform $Z_u \sim F_U$. Fix realizations $\{y_t\}$, $\{z_t\}$ and $\{z_{ut}\}$ from respective distributions. If $z_{ut} = F_z(z_t)$, then

$$F^{\dagger F_z}(y_t + z_t - 1|\Omega_t, \theta_0) = F^{\dagger F_U}(y_t + z_{ut} - 1|\Omega_t, \theta_0),$$

where $F^{\dagger F_z}$ stresses dependence of F^\dagger on F_z in (1), $F^{\dagger F_U}$ is as F^\dagger in (2), continued by uniform, and Ω_t denotes here realized past. Therefore, although a continued variable Y_t^\dagger and its distribution F^\dagger depends on F_z , $F^\dagger(Y_t^\dagger|\Omega_t, \theta_0)$ is not and we can always use uniform variables Z for continuation without affecting any properties of tests based on U_t .

Now we can use the fact that under the null hypothesis $U_t = F^\dagger(Y_t^\dagger|\Omega_t, \theta_0)$, $t = 1, \dots, T$, are uniform on $[0,1]$ and iid random variables, so that $P(U_{t-1} \leq r_1, U_{t-2} \leq r_2, \dots, U_{t-p} \leq r_p) = r_1 r_2 \dots r_p$, for $r = (r_1, \dots, r_p) \in [0,1]^p$. This motivates us to consider the following empirical processes

$$V_{pT}(r) = \frac{1}{\sqrt{T-(p+1)}} \sum_{t=p+1}^T \left[\prod_{j=1}^p I(U_{t-j} \leq r_j) - r_1 r_2 \dots r_p \right].$$

If we do not know θ_0 either $\{(Y_t, X_t), t \leq 0\}$, we approximate U_t with $\hat{U}_t = F_t^\dagger(Y_t^\dagger|\tilde{\Omega}_t, \hat{\theta})$ where $\hat{\theta}$ is an estimator of θ_0 and the truncated information set is $\tilde{\Omega}_t = \{X_t, X_{t-1}, \dots, X_1; Y_{t-1}, Y_{t-2}, \dots, Y_1\}$ and write

$$\hat{V}_{pT}(r) = \frac{1}{\sqrt{T-(p+1)}} \sum_{t=p+1}^T \left[\prod_{j=1}^p I(\hat{U}_{t-j} \leq r_j) - r_1 r_2 \dots r_p \right] \quad (3)$$

and

$$D_{pT} = \Gamma(\hat{V}_{pT}(r))$$

for any continuous functional $\Gamma(\cdot)$ from $\ell^\infty([0,1]^p)$, the set of uniformly bounded real functions on $[0,1]^p$, to R . In particular we use the Cramer-von Misses and Kolmogorov Smirnov test statistics

$$D_{pT}^{CvM} = \int_{[0,1]^p} \hat{V}_{pT}(r)^2 dr \text{ or } D_{pT}^{KS} = \max_{[0,1]^p} |\hat{V}_{pT}(r)|. \quad (4)$$

One further possibility is to test for j -lag pairwise independence, using the process

$$\hat{V}_{2T,j}(r) = \frac{1}{\sqrt{T-j}} \sum_{t=j+1}^T \left[I(\hat{U}_t \leq r_1) I(\hat{U}_{t-j} \leq r_2) - r_1 r_2 \right], \quad (5)$$

and corresponding test statistics $D_{2T,j}^{CvM}$ and $D_{2T,j}^{KS}$, say.

We can aggregate across p or j summing possibly with different weights $k(\cdot)$, obtaining generalized statistics

$$ADP_T = \sum_{p=1}^{T-1} k(p) D_{pT}, \text{ or } ADJ_T = \sum_{j=1}^{T-1} k(j) D_{2T,j}. \quad (6)$$

For $p = 1$, D_{1T}^{KS} delivers a generalization of Kolmogorov test to discrete distributions. Usually this test captures general deviations of marginal distribution but lacks power if only dynamics is misspecified. In particular, it does not have power against alternatives where U_t are uniform on $[0,1]$ but not independent. For general p , V_{pT} delivers a generalization of Kheifets (2011) to discrete

distributions. This test should capture both deviations of marginal distribution and deviations in dynamics.

A more direct approach is based in Box-Pierce (1970) type of statistics, we could consider

$$BPU_m := T \sum_{j=1}^m \hat{\rho}_{T,U}(j)^2,$$

$m = 1, 2, \dots$, and $\hat{\rho}_{T,U}(j)$ are the sample correlation coefficients of the U_t 's at lag j . Noting that U_t should be uniform continuous iid random variables under the null of correctly specified model, but might be correlated under alternative hypothesis of wrong specification, BPU_m is a good basis to design goodness-of-fit tests. This idea is related to the tests of Hong (1998). Alternatively, we can check autocorrelations of Gaussian residuals $\Phi(U_t)$

$$BPN_m := T \sum_{j=1}^m \hat{\rho}_{T,\Phi(U)}(j)^2,$$

and normality of $\Phi(U_t)$ with Jarque-Bera test. In addition we can check autocorrelations of discrete innovations,

$$e_t = \frac{Y_t - E[Y_t|\Omega_t]}{(\text{Var}[Y_t|\Omega_t])^{1/2}},$$

which are just the usual standardized probit residuals. We can define

$$BPD_m := T \sum_{j=1}^m \hat{\rho}_{T,e}(j)^2$$

and other statistics based on autocorrelations of squares of different types of residuals. The asymptotic distribution of these statistics can be approximated by chi square distributions when the true parameters θ_0 are known. Unlike tests based on empirical process, these tests can not capture some alternatives, for example if misspecification involves only higher order moments.

Parameter estimation affects the asymptotic distribution of these statistics, as well as that of those tests based on the empirical distribution of the U_t 's. There are different bootstrap and sampling techniques to approximate asymptotic distribution, see for example Shao and Dongsheng (1995), Politis, Romano and Wolf (1999). Since under H_0 we know the parametric conditional distribution, we apply parametric bootstrap to mimic the H_0 distribution. We introduce the algorithm now for statistics $\Gamma(\hat{V}_{2T})$.

1. Estimate model with initial data (Y_t, X_t) , $t = 1, 2, \dots, T$, get parameter estimator $\hat{\theta}$, get test statistic $\Gamma(\hat{V}_{2T})$.
2. Simulate Y_t^* with $F(\cdot|\Omega_t^*, \hat{\theta})$ recursively for $t = 1, 2, \dots, T$, where the bootstrap information set is $\Omega_t^* = (X_t, X_{t-1}, \dots, Y_{t-1}^*, Y_{t-2}^*, \dots)$.
3. Estimate model with simulated data Y_t^* , get θ^* , get bootstrapped statistics $\Gamma(\hat{V}_{2T}^*)$.
4. Repeat 2-3 B times, compute the percentiles of the empirical distribution of the B bootstrapped statistics.
5. Reject H_0 if $\Gamma(\hat{V}_{2T})$ is greater than the corresponding $(1 - \alpha)$ th percentile.

We will prove that $\Gamma(\hat{V}_{2T}^*)$ has the same limiting distribution as $\Gamma(\hat{V}_{2T})$. Bootstrapping other statistics is similar.

3 Asymptotic properties of specification tests

In this section we derive asymptotic properties of the statistics based on V_{2T} . We start with the simple case when we know parameters, then study how the asymptotic distribution changes if we estimate parameters. We provide analyses under the null, under the local and fixed alternatives. We first state all necessarily assumptions and propositions, then discuss them.

Let $\|\cdot\|$ denote Euclidean norm for matrices, i.e. $\|A\| = \sqrt{\text{tr}(A'A)}$ and for $\varepsilon > 0$, $B(a, \varepsilon)$ is an open ball in R^L with the center in the point a and the radius ε . In particular, for some $M > 0$ denote $B_T = B(\theta_0, MT^{-1/2}) = \{\theta : \|\theta - \theta_0\| \leq MT^{-1/2}\}$. For any discrete distributions G and F , with probability functions P_G and P_F , and $r \in [0, 1]$ define

$$\begin{aligned} d(G, F, r) &= G(F^{-1}(r)) - F(F^{-1}(r)) \\ &\quad + \frac{r - F(F^{-1}(r))}{P_F(F^{-1}(r) + 1)} (P_G(F^{-1}(r) + 1) - P_F(F^{-1}(r) + 1)). \end{aligned}$$

We have $d(F, F, r) = 0$, but $d(G, F, r)$ is not symmetric in G and F .

Assumption 1 Uniform boundedness away from zero: $\forall \varepsilon > 0, \exists \delta > 0$, such that $|F(0|\Omega_t, \theta)| > \varepsilon$ and $|F(j|\Omega_t, \theta) - F(j-1|\Omega_t, \theta)| > \varepsilon$ for $j = 1, \dots, J$ uniformly in $\theta \in B(\theta_0, \delta)$.

Assumption 2 Smoothness with respect to parameters:

(2.1)

$$E \max_{t=1, \dots, T} \sup_{u \in B_T} \max_y |F(y|\Omega_t, u) - F(y|\Omega_t, \theta_0)| = O(T^{-1/2}).$$

(2.2) $\forall M \in (0, \infty), \forall M_2 \in (0, \infty)$ and $\forall \delta > 0$

$$\max_y \frac{1}{\sqrt{T}} \sum_{t=1}^T \sup_{\substack{u, v \in B_T \\ \|u-v\| \leq M_2 T^{-1/2-\delta}}} |F(y|\Omega_t, u) - F(y|\Omega_t, v)| = o_p(1).$$

(2.3) $\forall M \in (0, \infty)$, there exists a uniformly continuous (vector) function $h(r)$ from $[0, 1]^2$ to R^L , such that

$$\sup_{v \in B_T} \sup_{r \in [0, 1]^2} \left| \frac{1}{\sqrt{T}} \sum_{t=2}^T h_t(r, v) - h(r)' \sqrt{T}(\theta_0 - v) \right| = o_p(1),$$

where

$$\begin{aligned} h_t(r, v) &= d(F(\cdot|\Omega_{t-1}, \theta_0), F(\cdot|\Omega_{t-1}, v), r_2) r_1 \\ &\quad + d(F(\cdot|\Omega_t, \theta_0), F(\cdot|\Omega_t, v), r_1) I(F(Y_{t-1}|\Omega_{t-1}, \theta_0) \leq r_2). \end{aligned}$$

Assumption 3 Linear expansion of the estimator: when the sample is generated by the null $F_t(y|\Omega_t, \theta_0)$, the estimator $\hat{\theta}$ admits a linear expansion

$$\sqrt{T}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{T}} \sum_{t=1}^T \ell(Y_t, \Omega_t) + o_p(1), \quad (7)$$

with $E_{F_t}(\ell(Y_t, \Omega_t) | \Omega_t) = 0$ and $\frac{1}{T} \sum_{t=1}^T \ell(Y_t, \Omega_t) \ell(Y_t, \Omega_t)' \xrightarrow{P_{F_t}} \Psi$.

Dynamic probit/logit and general discrete choice models considered in Introduction can easily be adjusted to satisfy all these assumptions. Discrete support allows a simple analytical closed form of conditional distribution of continued variable by any continuous random variable on unit support as in (2). Assumption 1 in particular requires that $F(0|\Omega_t, \theta)$ and $F(j|\Omega_t, \theta) - F(j-1|\Omega_t, \theta)$ for $j = 1, \dots, J$ are bounded away from zero uniformly around θ_0 . To study parameter estimation effect

we need to assume some smoothness of the distribution with respect to the parameter in Assumption 2 and a linear expansion of the estimator in Assumption 3. Note, the smoothness of the distribution with respect to the parameter is preserved after continuation, therefore Assumption 2 is similar to continuous case in Kheifets (2011); local Lipschitz continuity or existence of uniformly bounded first derivative of the distribution w.r.t. parameter is sufficient. For bootstrap we will need to strengthen Assumption 3 (see Assumption 3B below), although both conditions are standard and satisfied for many estimators, for example for MLE. Note, that to establish the convergence of the process V_{2T} (with known θ_0) under the null (the following Proposition 2), we do not need these assumptions.

We now describe the asymptotic behavior of the process $V_{2T}(r)$ under H_0 . Denote by " \Rightarrow " weak convergence of stochastic processes as random elements of the Skorokhod space $D([0, 1]^2)$.

Proposition 2 *Under H_0*

$$V_{2T} \Rightarrow V_{2\infty},$$

where $V_{2\infty}(r)$ is bi-parameter zero mean Gaussian process with covariance

$$\text{Cov}_{V_{2\infty}}(r, s) = (r_1 \wedge s_1)(r_2 \wedge s_2) + (r_1 \wedge s_2)r_2s_1 + (r_2 \wedge s_1)r_1s_2 - 3r_1r_2s_1s_2.$$

To take into account the estimation effect on the asymptotic distribution, we use a Taylor expansion to approximate $\hat{V}_{2T}(r)$ with $V_{2T}(r)$,

$$\hat{V}_{2T}(r) = V_{2T}(r) + \sqrt{T} \left(\hat{\theta} - \theta_0 \right)' h(r) + o_p(1)$$

uniformly in r . To identify the limit of $\hat{V}_{2T}(r)$, we need to study limiting distribution of $\sqrt{T}(\hat{\theta} - \theta_0)$, using the expansion from Assumption 3. Define

$$C_T(r, s, \theta) = E \left(\begin{pmatrix} V_{2T}(r) \\ \frac{1}{\sqrt{T}} \sum_{t=1}^T \ell(Y_t, \Omega_t) \end{pmatrix} \begin{pmatrix} V_{2T}(s) \\ \frac{1}{\sqrt{T}} \sum_{t=1}^T \ell(Y_t, \Omega_t) \end{pmatrix}' \right)$$

and let $(V_{2\infty}(r), \psi'_\infty)'$ be a zero mean Gaussian process with covariance function $C(r, s, \theta_0) = \lim_{T \rightarrow \infty} C_T(r, s, \theta_0)$. Dependence on θ on right hand side (rhs) comes through U_t and $\ell(\cdot, \cdot)$.

Suppose the conditional distribution function $H(y|\Omega_t)$ is not in the parametric family $F(y|\Omega_t, \theta)$ but has the same support. For any $T_0 \in \{0, 1, 2, \dots\}$ and $T \geq T_0$ define conditional on Ω_t conditional df

$$G_T(y|\Omega_t, \theta) = \left(1 - \frac{\sqrt{T_0}}{\sqrt{T}} \right) F(y|\Omega_t, \theta) + \frac{\sqrt{T_0}}{\sqrt{T}} H(y|\Omega_t).$$

Now we define local alternatives:

H_{1T} : Conditional cdf of Y_t is equal to $G_T(y|\Omega_t, \theta_0)$ with $T_0 \neq 0$.

Conditional cdf $G_T(y|\Omega_t, \theta_0)$ allow us to study all three cases: H_0 if $T_0 = 0$, H_{1T} if $T = T_0, T_0 + 1, T_0 + 2, \dots$ and $T_0 \neq 0$ and H_1 if we fix $T = T_0$. In the next proposition we provide the asymptotic distribution of our statistics under the null and under the local alternatives.

Proposition 3 *a) Suppose Assumptions 1-3 hold. Then under H_0*

$$\Gamma(\hat{V}_{2T}) \xrightarrow{d} \Gamma(\hat{V}_{2\infty}),$$

where $\hat{V}_{2\infty}(r) = V_{2\infty}(r) - h(r)' \psi_\infty$.

b) Suppose Assumptions 1-3 hold. Then under H_{1T}

$$\Gamma(\hat{V}_{2T}) \xrightarrow{d} \Gamma \left(\hat{V}_{2\infty} + \sqrt{T_0}k - \sqrt{T_0}\xi'h \right),$$

where

$$k(r) = \text{plim}_{T \rightarrow \infty} \frac{1}{T} \sum_{t=2}^T \{d(H(\cdot|\Omega_{t-1}), F(\cdot|\Omega_{t-1}, \theta_0), r_2) r_1 + d(H(\cdot|\Omega_t), F(\cdot|\Omega_t, \theta_0), r_1) I(F(Y_{t-1}|\Omega_{t-1}, \theta_0) \leq r_2)\},$$

and

$$\xi = \text{plim}_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \ell(Y_t, \Omega_t). \quad (8)$$

Under G_T , the random variables $U_t = F^\dagger(Y_t^\dagger|\Omega_t, \theta_0)$ are not anymore iid, instead $U_t^* = G_T^\dagger(Y_t^\dagger|\Omega_t, \theta_0)$ are uniform iid. The first term in $k(r)$ controls for the lack of uniformity of U_t (and it is similar to Bai's (2003) $k(r)$), it is zero when U_t are uniform. The second term in $k(r)$ adds control for independence of U_t , cf. Kheifets (2011).

Under the alternative we may have also that (7) is not centred around zero, since $E_{G_T}(\ell(Y_t, \Omega_t)|\Omega_t) = \frac{\sqrt{T_0}}{\sqrt{T}} E_H(\ell(Y_t, \Omega_t)|\Omega_t)$, therefore ξ may be nonzero, which stands for information from estimation. This term does not appear in Bai (2003) method, since his method projects out the estimation effect.

For the case of the one parameter empirical process, we can provide the following corollary, which is similar to Bai (2003)'s single parameter results.

Corollary 4 *a) Suppose Assumptions 1-3 hold. Then under H_0*

$$\Gamma(\hat{V}_{1T}(\cdot)) \xrightarrow{d} \Gamma(\hat{V}_{2\infty}(\cdot, 1)),$$

where $\hat{V}_{1\infty}(\cdot) = V_{1\infty}(\cdot) - h(\cdot, 1)' \psi_\infty$ and $V_{1\infty}(\cdot) = V_{2\infty}(\cdot, 1)$.

b) Suppose Assumptions 1-3 hold. Then under H_{1T}

$$\Gamma(\hat{V}_{1T}(\cdot)) \xrightarrow{d} \Gamma(\hat{V}_{1\infty}(\cdot) + \sqrt{T_0} k_1(\cdot) - \sqrt{T_0} h(\cdot, 1)' \xi),$$

where for $r \in [0, 1]$

$$k_1(r) = \text{plim}_{T \rightarrow \infty} \frac{1}{T} \sum_{t=2}^T d(H(\cdot|\Omega_t), F(\cdot|\Omega_t, \theta_0), r).$$

Note then that tests based on \hat{V}_{1T} are not consistent against alternatives for which $k_1 = 0$ and $h(\cdot, 1) = 0$ but $k \neq 0$ or $h(\cdot, 1) \neq 0$ on some set of positive measure.

We will justify our bootstrap procedure now, i.e. we prove that $\Gamma(\hat{V}_{2T}^*)$ has the same limiting distribution as $\Gamma(\hat{V}_{2T})$. We say that the sample is distributed under $\{\theta_T : T \geq 1\}$ when there is a triangular array of random variables $\{Y_{Tt} : T \geq 1, t \leq T\}$ with (T, t) element generated by $F(\cdot|\Omega_{Tt}, \theta_T)$, where $\Omega_{Tt} = (X_{t-1}, X_{t-2}, \dots, Y_{Tt-1}, Y_{Tt-2}, \dots)$. Similar arguments can be applied to other statistics.

Assumption 3B For all nonrandom sequences $\{\theta_T : T \geq 1\}$ for which $\theta_T \rightarrow \theta_0$, we have

$$\sqrt{T}(\hat{\theta} - \theta_T) = \frac{1}{\sqrt{T}} \sum_{t=1}^T \ell(Y_{Tt}, \Omega_{Tt}) + o_p(1),$$

under $\{\theta_T : T \geq 1\}$, where $E[\ell(Y_{Tt}, \Omega_{Tt})|\Omega_{Tt}] = 0$ and

$$\frac{1}{T} \sum_{t=1}^T \ell(Y_{Tt}, \Omega_{Tt}) \ell(Y_{Tt}, \Omega_{Tt})' \xrightarrow{p} \Psi.$$

Note that the function $\ell(\cdot, \cdot)$ now depends on θ_T and is assumed to be the same as in Assumption 3. We require that estimators of close to θ_0 points have *the same* linear representation as the estimator of θ_0 itself.

Table 1: Different scenarios for Monte Carlo experiments.

	DGP	Null
1	probit static	probit static
2	probit dynamic	probit dynamic
3	probit interactions	probit interactions
4	logit static	probit static
5	chi2 static	probit static
6	logit dynamic	probit static
7	chi2 dynamic	probit static
8	logit interactions	probit dynamic
9	chi2 interactions	probit dynamic
10	logit interactions	probit static
11	chi2 interactions	probit static

Proposition 5 *Suppose Assumptions 1, 2 and 3B hold. Then for any nonrandom sequence $\{\theta_T : T \geq 1\}$ for which $\theta_T \rightarrow \theta_0$, under $\{\theta_T : T \geq 1\}$,*

$$\Gamma(\hat{V}_{2T}(r)) \xrightarrow{d} \Gamma(\hat{V}_{2\infty}(r)).$$

4 Monte Carlo Simulation

In this section we investigate the finite sample properties of our bootstrap tests using Monte Carlo exercise. We use a simple dynamic Probit model with one exogenous regressor with autoregressive dynamics. We consider three specifications of dynamics

$$\begin{aligned} \text{Static model} &: \pi_t = \pi_0 + \beta X_t, \\ \text{Dynamic model} &: \pi_t = \pi_0 + \delta_1 Y_{t-1} + \beta X_t, \\ \text{Dynamic model with interactions} &: \pi_t = \pi_0 + \delta_1 Y_{t-1} + \gamma_1 Y_{t-1} X_t + \beta X_t, \quad \gamma_1 = -2\beta, \end{aligned}$$

where in all specifications X_t follows an AR(1) process,

$$X_t = \alpha_1 X_{t-1} + e_t, \quad e_t \sim IIN(0, 1),$$

and we set $\pi_0 = 0, \beta = 1, \delta_1 = 0.8, \alpha_1 = 0.8$.

We try 11 different scenarios of data generating processes (DGP) and null hypotheses (see Table 1). In the first three we study the size properties of static, dynamic and dynamic with interactions probit models. Other scenarios check power when dynamics and/or marginals are misspecified. We take logit and $(\chi_1^2 - 1)/2^{1/2}$ as alternative distributions. We use sample sizes $T = 100$ (Table 2), 300 (Table 3) and 500 (Table 4) with 1000 replications. To estimate the Bootstrap percentages of rejections we use a Warp bootstrap Monte Carlo (see Giacomini, Politis and White, 2007) for all considered test statistics. For tests based on "continued" residuals we consider one-parameter ($p = 1$) and two-parameter empirical processes ($p = 2$) with $j = 1$ and $j = 2$ lags and Cramer-von Misses (CvM) and Kolmogorov-Smirnov (KS) criterions. To make the results more readable, we denote them as $CvM_0 = D_{1T}^{CvM}$, $CvM_1 = D_{2T,1}^{CvM}$, $CvM_2 = D_{2T,2}^{CvM}$ and $KS_0 = D_{1T}^{KS}$, $KS_1 = D_{2T,1}^{KS}$, $KS_2 = D_{2T,2}^{KS}$. We consider Box-Pierce type tests for Gaussian and discrete residuals with $m = 1, 2, 25$. We also check normality of Gaussian residuals with a bootstrapped Jarque-Bera test (JB). The results of empirical process tests with further lags $j = 3, 4, 5$ and correlation tests on uniform residuals do not provide additional information and are omitted.

Table 2: Percentage of rejections of test statistics with $T = 100$.

		CvM_0	CvM_1	CvM_2	KS_0	KS_1	KS_2	BPN_1	BPN_2	BPN_{25}	JB	BPD_1	BPD_2	BPD_{25}
1	10%	8.8	7.4	10.4	8.4	10.1	9.2	9.5	9.6	9.3	8.3	10.1	10.6	8.8
	5%	3.5	4.3	4.3	3.9	4.8	4.7	4.6	3.7	3.8	4.4	5.5	5.1	3.4
	1%	0.3	0.9	0.4	0.5	1.1	1.7	1.5	0.8	0.3	2.1	0.8	0.6	0.3
2	10%	7.9	8.3	8.7	7.0	9.6	9.2	9.0	10.6	7.0	11.2	9.5	10.7	12.8
	5%	3.0	3.6	4.0	2.8	4.9	4.5	6.0	4.0	2.1	5.9	4.0	4.4	6.1
	1%	0.0	0.4	0.1	0.8	1.2	1.0	0.9	1.3	0.3	1.5	0.4	1.1	0.7
3	10%	8.9	10.0	9.5	7.7	10.6	9.4	10.1	11.3	8.9	10.7	9.2	9.2	10.1
	5%	4.1	4.1	3.9	3.6	4.9	5.0	5.5	5.5	4.5	5.4	5.5	3.8	5.4
	1%	0.1	0.1	0.2	1.1	0.5	0.8	1.2	1.1	0.5	1.1	0.6	0.8	0.5
4	10%	8.1	9.0	7.6	8.4	8.9	9.9	7.2	8.8	7.5	9.9	9.0	9.2	9.0
	5%	3.9	4.6	3.5	3.6	4.1	3.7	3.5	4.1	3.6	3.0	5.1	4.6	4.1
	1%	0.5	0.4	0.3	0.6	0.6	0.6	1.2	0.7	0.6	0.5	1.0	1.9	0.7
5	10%	10.4	9.5	10.2	12.0	10.1	11.1	9.2	11.5	10.7	20.3	8.0	7.5	9.2
	5%	4.9	6.1	5.6	5.9	5.2	5.7	5.7	6.3	6.1	12.6	4.6	3.7	4.8
	1%	0.5	0.9	0.3	0.8	1.2	0.3	1.0	1.0	0.7	4.3	1.8	1.6	0.9
6	10%	9.5	11.0	7.6	9.2	9.8	9.3	19.1	15.4	11.7	11.0	43.0	35.3	16.8
	5%	4.6	4.9	3.5	3.5	5.2	4.6	10.7	9.0	6.6	4.7	29.4	20.5	9.4
	1%	0.4	0.5	0.8	0.5	1.4	0.9	2.9	2.3	0.8	0.9	11.0	5.7	2.9
7	10%	10.3	10.9	9.4	9.2	10.0	9.3	28.3	26.4	14.5	13.7	60.0	50.6	24.6
	5%	4.8	5.2	4.7	3.9	4.7	5.2	20.6	16.4	8.5	7.7	47.1	37.0	16.7
	1%	0.1	1.5	0.1	1.2	1.3	0.5	9.4	6.0	2.6	2.3	26.0	16.4	5.6
8	10%	9.7	9.2	13.7	9.9	10.1	13.0	14.0	26.6	16.2	9.9	46.2	57.4	30.1
	5%	4.0	3.7	7.8	3.6	5.5	7.8	6.5	18.8	10.1	3.8	36.9	45.1	18.6
	1%	0.8	0.8	2.5	0.8	1.1	1.3	0.6	6.4	2.3	0.3	17.1	27.6	5.0
9	10%	14.4	16.9	29.1	16.0	20.6	34.4	18.1	55.7	33.5	20.0	79.0	82.2	64.9
	5%	8.9	10.0	21.1	9.9	12.5	18.5	11.2	48.4	23.1	12.4	72.9	81.0	59.8
	1%	0.9	1.8	3.9	1.4	4.2	3.8	3.4	26.9	11.6	3.1	58.1	77.2	43.8
10	10%	8.6	14.7	18.1	7.6	15.5	13.0	28.0	42.0	21.3	9.2	50.1	79.8	43.6
	5%	2.8	8.5	9.5	3.8	7.6	6.6	17.5	29.0	12.9	3.9	35.7	69.8	30.4
	1%	0.6	1.4	2.1	0.4	1.3	0.5	5.3	12.4	4.4	0.3	22.4	45.6	10.7
11	10%	9.0	28.1	33.6	8.1	29.2	28.4	53.1	85.1	60.3	8.8	72.0	99.9	94.2
	5%	3.4	17.6	19.8	3.3	17.1	11.8	40.7	76.3	43.1	5.3	61.6	99.7	90.5
	1%	0.2	6.1	4.2	0.2	3.4	1.4	23.2	57.1	22.0	0.6	40.8	98.4	72.3

Table 3: Percentage of rejections of test statistics with $T = 300$.

		CvM_0	CvM_1	CvM_2	KS_0	KS_1	KS_2	BPN_1	BPN_2	BPN_{25}	JB	BPD_1	BPD_2	BPD_{25}
1	10%	8.3	9.2	9.2	9.1	9.0	10.4	8.0	8.3	8.6	8.0	8.3	8.2	10.2
	5%	4.1	4.7	4.6	5.2	4.8	4.6	3.6	3.9	4.4	2.7	4.2	3.1	4.8
	1%	0.7	1.0	0.6	0.8	0.6	0.4	0.6	0.7	1.1	0.6	0.6	0.6	1.0
2	10%	9.3	8.9	10.2	9.3	9.8	11.6	7.6	10.0	9.7	9.6	9.0	6.9	7.4
	5%	5.5	4.4	5.9	5.2	4.9	6.4	3.8	4.6	4.8	3.8	3.7	3.6	3.2
	1%	1.0	1.1	0.9	1.4	1.1	0.9	0.5	0.7	1.4	0.7	0.5	0.2	0.6
3	10%	8.5	12.3	8.7	8.7	12.2	9.9	8.1	9.9	10.1	9.0	10.1	9.4	13.5
	5%	4.5	5.1	5.1	3.3	5.4	5.3	4.4	5.2	5.9	4.2	5.3	4.2	5.1
	1%	1.1	1.0	1.5	1.2	0.6	0.9	0.9	1.0	1.3	0.5	0.6	1.2	1.5
4	10%	9.9	10.2	9.2	9.5	10.8	9.8	8.6	9.3	10.0	11.2	8.5	9.6	8.5
	5%	4.8	4.9	4.6	5.3	5.2	4.9	4.4	5.1	4.0	6.1	3.3	3.2	3.6
	1%	0.9	1.1	0.6	0.9	1.3	1.1	0.5	0.5	1.0	0.7	0.8	0.5	0.9
5	10%	16.5	15.1	14.9	15.8	14.9	14.4	11.8	11.2	8.7	41.4	6.0	7.2	9.0
	5%	8.8	7.9	8.3	9.6	7.5	8.5	5.0	6.1	4.1	30.4	4.6	6.3	6.7
	1%	1.6	2.0	1.8	1.8	2.2	1.4	0.9	1.4	1.0	9.6	3.0	3.1	2.9
6	10%	8.8	15.4	11.3	9.0	13.7	10.1	38.3	29.9	16.0	9.6	79.1	69.0	29.2
	5%	4.7	9.8	4.8	5.4	7.9	6.1	25.1	21.8	8.5	5.8	65.2	58.1	19.3
	1%	0.5	2.0	0.3	0.6	1.5	1.5	11.8	5.8	1.5	0.7	43.7	36.6	6.8
7	10%	11.8	12.2	14.4	12.3	8.8	13.2	42.5	35.1	15.6	24.5	55.6	47.1	24.0
	5%	7.0	5.9	9.4	6.2	4.1	6.3	31.2	25.0	9.9	16.3	44.6	39.5	15.0
	1%	1.1	1.3	1.9	1.2	1.2	2.0	15.3	9.2	2.6	3.3	35.3	20.2	5.8
8	10%	12.5	18.3	44.9	13.0	17.2	44.0	19.5	67.0	30.0	12.6	91.4	96.5	72.0
	5%	6.3	12.5	31.6	6.4	10.3	28.9	10.6	55.7	19.3	6.2	85.6	94.0	59.5
	1%	1.0	2.6	9.8	0.9	2.7	8.5	2.7	31.3	7.4	1.4	71.4	87.7	37.3
9	10%	32.9	42.5	81.0	29.5	46.9	88.8	34.3	92.0	71.3	24.8	99.0	99.7	98.8
	5%	17.3	25.9	65.2	19.7	36.3	80.2	25.2	88.8	64.7	17.6	98.0	99.6	97.8
	1%	3.0	7.8	36.3	3.9	14.3	56.4	12.3	77.1	42.1	4.0	94.5	99.1	95.6
10	10%	8.7	33.1	44.9	9.7	28.3	33.7	51.9	83.2	49.4	9.3	81.7	99.6	84.6
	5%	4.4	22.4	30.0	4.7	17.7	21.4	36.8	74.5	35.2	5.5	69.5	98.8	78.7
	1%	1.1	9.0	11.1	0.8	6.8	5.2	18.2	54.0	17.0	1.2	37.9	97.1	48.1
11	10%	8.6	46.2	76.7	9.1	46.7	76.9	63.8	99.5	89.7	11.0	81.5	100.0	100.0
	5%	4.3	32.8	63.9	4.3	36.1	67.8	51.0	98.8	81.8	5.6	68.2	100.0	100.0
	1%	0.4	11.7	37.3	0.4	18.4	42.8	31.7	94.9	63.9	0.9	39.4	100.0	99.3

Table 4: Percentage of rejections of test statistics with $T = 500$.

		CvM_0	CvM_1	CvM_2	KS_0	KS_1	KS_2	BPN_1	BPN_2	BPN_{25}	JB	BPD_1	BPD_2	BPD_{25}
1	10%	10.2	8.1	9.2	9.8	8.6	8.0	11.9	11.7	11.1	11.3	10.1	8.6	9.8
	5%	4.7	4.5	4.9	4.2	3.9	4.3	6.0	5.4	5.6	5.3	5.2	4.7	4.7
	1%	0.5	1.0	0.7	0.7	0.6	1.0	0.6	0.8	0.8	1.2	1.1	0.4	1.1
2	10%	9.1	8.0	8.3	10.2	8.6	10.7	11.6	11.5	8.4	9.2	8.6	9.9	10.6
	5%	4.3	4.9	4.4	4.7	3.8	4.1	5.6	4.8	3.9	4.6	4.3	5.7	5.6
	1%	0.7	0.8	0.8	0.5	0.2	0.6	0.6	1.4	0.5	0.5	1.4	0.8	0.8
3	10%	9.1	8.9	9.9	9.2	8.6	8.5	10.0	11.6	10.7	11.0	10.7	10.2	10.0
	5%	4.1	3.9	3.1	4.8	4.5	4.9	5.7	7.1	5.4	5.2	4.6	5.3	4.8
	1%	0.7	1.1	1.0	1.0	0.4	0.4	1.1	1.0	1.9	1.5	1.5	2.0	1.6
4	10%	10.7	8.9	10.9	10.7	10.8	8.8	11.5	9.2	10.2	7.8	10.0	10.4	12.3
	5%	5.2	4.3	4.8	5.2	3.9	4.1	6.1	3.5	5.2	4.4	5.1	6.7	7.2
	1%	0.4	0.8	1.0	0.7	0.6	0.4	0.7	1.1	0.7	0.7	1.4	1.3	1.9
5	10%	17.1	16.8	17.0	20.0	19.4	17.8	11.8	12.1	8.5	53.2	7.9	8.0	14.0
	5%	11.4	10.0	9.7	13.4	12.0	12.1	4.4	5.5	3.8	45.1	5.6	5.4	9.1
	1%	3.6	3.7	3.2	4.5	4.9	3.3	1.3	1.7	0.7	23.6	3.4	3.7	5.3
6	10%	8.7	17.1	9.8	10.2	15.5	8.9	46.1	36.2	17.6	9.3	88.7	81.2	42.0
	5%	5.3	8.9	4.9	4.9	6.9	3.4	32.9	27.3	11.8	3.7	82.9	69.0	29.5
	1%	0.6	1.5	1.0	0.9	1.1	0.8	17.2	8.9	2.4	0.6	53.0	46.8	8.5
7	10%	15.2	11.9	14.8	13.3	11.3	15.9	39.9	36.3	18.0	38.0	53.7	42.6	21.5
	5%	8.1	5.6	10.8	7.5	4.8	8.0	28.5	28.0	10.6	27.9	41.0	34.4	16.5
	1%	2.2	1.3	2.8	2.5	1.7	3.0	12.0	9.4	3.0	8.2	18.2	19.5	9.4
8	10%	22.6	34.0	90.1	25.1	35.9	92.9	23.7	97.6	76.7	10.0	99.9	100.0	99.6
	5%	12.6	23.3	83.8	14.7	25.9	86.5	16.7	95.9	65.1	5.0	99.7	100.0	99.2
	1%	3.4	7.6	53.9	4.0	10.4	65.0	5.1	87.6	44.7	1.1	99.2	99.9	97.1
9	10%	56.1	73.2	98.8	58.3	78.1	99.6	62.6	99.5	96.9	30.4	100.0	100.0	100.0
	5%	39.9	59.9	97.0	41.9	69.9	98.8	51.2	99.3	95.4	20.5	100.0	100.0	100.0
	1%	13.4	38.7	88.5	16.1	48.2	95.6	31.1	98.6	91.4	10.7	100.0	100.0	99.8
10	10%	9.9	58.4	90.5	10.3	52.8	88.7	74.8	99.6	93.6	7.9	98.7	100.0	100.0
	5%	5.2	43.1	82.2	5.0	38.2	79.8	65.2	99.3	89.4	4.5	95.9	100.0	100.0
	1%	1.1	14.3	59.4	0.6	14.1	39.6	44.5	97.3	78.0	1.3	86.0	100.0	99.7
11	10%	10.0	74.7	99.1	11.3	81.3	98.0	86.8	100.0	100.0	9.4	99.6	100.0	100.0
	5%	3.6	62.5	96.7	4.8	71.3	95.6	78.9	100.0	99.9	4.6	98.6	100.0	100.0
	1%	0.5	38.5	87.5	0.8	33.1	80.7	64.0	100.0	99.9	1.3	87.1	100.0	100.0

Now we discuss the performance of empirical process based tests in comparison with traditional correlation tests. For $T = 100$ almost all test statistics are slightly undersized (cases 1-3). The situation improves with larger T , and CvM statistics approach faster to nominal rates than KS. Overall, empirical size at $T = 500$ is very good. The situation with power is not unambiguous. No test can capture static logit alternative to the null hypothesis of static probit model even at $T = 500$ (case 4). On the other hand, when static χ^2 alternative to the null hypothesis of static probit is considered (case 5), there is some power at $T = 300$ which improves with $T = 500$ for all empirical process based tests. Since under the null and under the alternative we have static models, correlation tests do not have power. Normality test (JB) is doing well only in the latter case. When there is a slight dynamic misspecification added to logit (case 6), CvM_1 and KS_1 improve, but when it is added to χ^2 our tests and JB doing worse (case 7). Correlation tests, on the contrary display power against these dynamic alternatives. When the alternative has dynamic interactions, and the null is a dynamic probit (cases 8 and 9), all tests (but JB for logit) are doing well, and even better if higher lags are taken into account. Finally, when dynamic interactions are taken versus static model (cases 10 and 11), power is very good, and increases when more lags are considered. Exceptions are "marginal tests" CvM_0 , KS_0 and JB. To summarize, dynamic misspecification can be captured well by empirical process statistics and correlation tests. Misspecification in marginals, on the contrary, can not be distinguished at all by correlation tests but empirical process statistics, possibly multi-parameter, still work, although further research in improving power of these tests is needed.

To develop our omnibus type tests we introduce additional continuous noise. An important question is the effect of this noise on the power of the tests. Since correlation tests based on discrete residuals BPD_j do not use additional noise, while correlation tests based on continuous residuals BPN_j do, we can use the difference in rejection rates between these sets of statistics under dynamic misspecification as an indirect measure of the effect of the introduced noise, though correlation tests are not consistent against static alternatives. From our Monte Carlo simulations we see that for all scenarios we consider, correlation tests based on discrete residuals perform better, indicating that some power losses may indeed be attributed to the introduced noise. To overcome this problem, we plan to develop tests for discrete models based on alternative transformations of the data without introducing additional noise, but still consistent against a wide range of nonparametric alternative hypotheses.

5 Conclusion

In this paper we have proposed new tests for checking goodness-of-fit of conditional distributions in nonlinear discrete time series models. Specification of the conditional distribution (but not only conditional moments) is important in many macroeconomics and financial applications. Due to the parameter estimation effect, the asymptotic distribution depends on the model and specific parameter values. We show that our parametric bootstrap provides a good approximation to asymptotic distributions and renders feasible and simple tests. Monte Carlo experiments have shown that tests based on empirical processes have power if misspecification comes from dynamics. If misspecification affects marginals alone, correlation tests are inconsistent, while tests based on empirical processes have some power. Comparing to the continuous case, we may conclude that there is a reduction of power due to the additional noise which distribution is known under the alternative too.

Appendix

Proof of Proposition 1. Part (a) is a property of dynamic PIT with a continuous conditional distribution F_t^\dagger , the proof can be found in Bai (2003). Part (b) follows from the fact that (omitting dependence on t , Ω_t and θ)

$$\begin{aligned} F^\dagger(Y + Z - 1) &= F([Y + Z - 1]) + Z^U \mathbf{P}([Y + Z]) \\ &= F(Y - 1) + Z^U \mathbf{P}(Y), \end{aligned}$$

where

$$Z^U = F_z(Y + Z - 1 - [Y + Z - 1]) = F_z(Z)$$

is uniform for any $Z \sim F_z$ continuous and with $[0, 1]$ support, by the usual static PIT property. Therefore, although a continued variable Y^\dagger and its distribution F^\dagger depends on F_z , $F^\dagger(Y^\dagger)$ does not. \square

Proof of Propositions 2. Assumption 1 in Kheifets (2011) is satisfied automatically after applying continuation defined in (2), therefore Proposition 1 of Kheifets (2011) holds. \square

Proof of Propositions 3. Follows from Kheifets (2011), we need only to check that Assumption 2 in Kheifets (2011) is satisfied.

Let $r = F^\dagger(y)$. Note that $[y] = F^{-1}(r)$ but $F([y]) = F(F^{-1}(r))$ equals r only when $y = [y]$. The inverse of F^\dagger is

$$\begin{aligned} y &= \left(F^\dagger\right)^{-1}(r) = [y] + \frac{r - F([y])}{\mathbf{P}([y] + 1)} = [y] + 1 + \frac{r - F([y] + 1)}{\mathbf{P}([y] + 1)} \\ &= F^{-1}(r) + \frac{r - F(F^{-1}(r))}{\mathbf{P}(F^{-1}(r) + 1)}. \end{aligned}$$

Note also that $(r - F([y])) / \mathbf{P}([y] + 1) = y - [y] \in [0, 1]$. Take distribution G with the same support as F . We have different useful ways to write $d(G, F, r)$:

$$\begin{aligned} d(G, F, r) &= \eta^\dagger(r) - r = G^\dagger\left(\left(F^\dagger\right)^{-1}(r)\right) - r = G^\dagger(y) - r \\ &= G([y]) - F([y]) + (y - [y])(\mathbf{P}_G([y] + 1) - \mathbf{P}_F([y] + 1)) \end{aligned} \quad (9)$$

$$\begin{aligned} &= G([y] + 1) - F([y] + 1) \\ &\quad + (y - [y] - 1)(\mathbf{P}_G([y] + 1) - \mathbf{P}_F([y] + 1)) \end{aligned} \quad (10)$$

$$\begin{aligned} &= G(F^{-1}(r)) - F(F^{-1}(r)) \\ &\quad + \frac{r - F(F^{-1}(r))}{\mathbf{P}_F(F^{-1}(r) + 1)} (\mathbf{P}_G(F^{-1}(r) + 1) - \mathbf{P}_F(F^{-1}(r) + 1)). \end{aligned} \quad (11)$$

Thus, noting that $\mathbf{P}(\cdot)$ is bounded away from zero, we have that Assumption 2 in this paper is sufficient for Assumption 2 in Kheifets (2011):

(K2.1)

$$E \sup_{t=1, \dots, T} \sup_{u \in B_T} \sup_{r \in [0, 1]} \left| \eta_t^\dagger(r, u, \theta_0) - r \right| = O\left(T^{-1/2}\right).$$

(K2.2) $\forall M \in (0, \infty)$, $\forall M_2 \in (0, \infty)$ and $\forall \delta > 0$

$$\sup_{r \in [0, 1]} \frac{1}{\sqrt{T}} \sum_{t=1}^T \sup_{\substack{\|u-v\| \leq M_2 T^{-1/2-\delta} \\ u, v \in B_T}} \left| \eta_t^\dagger(r, u, \theta_0) - \eta_t^\dagger(r, v, \theta_0) \right| = o_p(1).$$

(K2.3) $\forall M \in (0, \infty), \forall M_2 \in (0, \infty)$ and $\forall \delta > 0$

$$\sup_{|r-s| \leq M_2 T^{-1/2-\delta}} \frac{1}{\sqrt{T}} \sum_{t=1}^T \sup_{u \in B_T} \left| \eta_t^\dagger(r, u, \theta_0) - \eta_t^\dagger(s, u, \theta_0) \right| = o_p(1).$$

(K2.4) $\forall M \in (0, \infty)$, there exists a uniformly continuous (vector) function $h(r)$ from $[0, 1]^2$ to R^L , such that

$$\sup_{u \in B_T} \sup_{r \in [0, 1]^2} \left| \frac{1}{\sqrt{T}} \sum_{t=2}^T h_t - h(r)' \sqrt{T} (u - \theta_0) \right| = o_p(1).$$

where

$$h_t = \left(\eta_{t-1}^\dagger(r_2, u, \theta_0) - r_2 \right) r_1 + \left(\eta_t^\dagger(r_1, u, \theta_0) - r_1 \right) I \left(F_{t-1}^\dagger \left(Y_{t-1}^\dagger | u \right) \leq r_2 \right).$$

For Part a), take $d \left(F(\cdot | \Omega_t, \theta_0), F(\cdot | \Omega_t, \hat{\theta}) \right)$. Then (K2.1), (K2.2), (K2.4) follow from (2.1), (2.2) and (2.3) because of representation (11). If we compare (9) and (10) we see that $d(\cdot)$ is not only continuous in r , but piece-wise linear, so (K2.3) is satisfied automatically.

For Part b), take $d \left(G_T(\cdot | \Omega_t, \theta_0), F(\cdot | \Omega_t, \hat{\theta}) \right)$ and use the additivity of $d(\cdot)$ in the first arguments:

$$\begin{aligned} d \left(G_T(\cdot | \Omega_t, \theta_0), F(\cdot | \Omega_t, \hat{\theta}) \right) &= \left(1 - \frac{\sqrt{T_0}}{\sqrt{T}} \right) d \left(F(\cdot | \Omega_t, \theta_0), F(\cdot | \Omega_t, \hat{\theta}) \right) \\ &\quad + \frac{\sqrt{T_0}}{\sqrt{T}} d \left(H(\cdot | \Omega_t), F(\cdot | \Omega_t, \hat{\theta}) \right). \end{aligned}$$

Proof of Propositions 5. The proof is similar if we consider $d \left(F(\cdot | \Omega_t, \theta_T), F(\cdot | \Omega_t, \hat{\theta}_T) \right)$ under $\{\theta_T : T \geq 1\}$. □

References

- [1] Andrews, D.W.K. (1997). A conditional Kolmogorov test. *Econometrica* 65, 1097- 1128.
- [2] Bai, J. (2003). Testing Parametric Conditional Distributions of Dynamic Models. *Review of Economics and Statistics* 85, 531-549.
- [3] Bai, J. and S. Ng (2001). A consistent test for conditional symmetry in time series models. *Journal of Econometrics* 103, 225-258.
- [4] Basu, D. and R. de Jong (2007). Dynamic Multinomial Ordered Choice with an Application to the Estimation of Monetary Policy Rules. *Studies in Nonlinear Dynamics and Econometrics* 4, article 2.
- [5] Blum, J. R., Kiefer, J. and M. Rosenblatt (1961). Distribution free tests of independence based on sample distribution function. *Annals of Mathematical Statistics* 32, 485-98.
- [6] Box, G. and D. Pierce (1970). Distribution of residual autocorrelations in autorregressive integrated moving average time series models. *Journal of the American Statistical Association* 65, 1509-1527.
- [7] Corradi, V. and R. Swanson (2006). Bootstrap conditional distribution test in the presence of dynamic misspecification. *Journal of Econometrics* 133, 779-806.
- [8] de Jong, R.M. and T. Woutersen (2011). Dynamic time series binary choice. *Econometric Theory* 27, 673-702.
- [9] Delgado, M. (1996). Testing serial independence using the sample distribution function, *Journal of Time Series Analysis* 17, 271-285.
- [10] Delgado, M. and J. Mora (2000). A nonparametric test for serial independence of regression errors, *Biometrika* 87, 228-234.
- [11] Delgado, M. and W. Stute (2008). Distribution-free specification tests of conditional models. *Journal of Econometrics* 143, 37-55.
- [12] Denuit, M. and P. Lambert (2005). Constraints on concordance measures in bivariate discrete data. *Journal of Multivariate Analysis* 93, 40-57.
- [13] Dueker, M. (1997). Strengthening the case for the yield curve as a predictor of U.S. recessions. Review, Federal Reserve Bank of St. Louis, issue Mar, 41-51.
- [14] Ferguson, T.S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press.
- [15] Giacomini, R., D.N. Politis, and H. White (2007). A Warp-Speed Method for Conducting Monte Carlo Experiments Involving Bootstrap Estimators. Mimeo.
- [16] Hamilton, J. and O. Jorda (2002). A model of the Federal Funds rate target. *Journal of Political Economy* 110, 1135-1167.
- [17] Hoeffding W. (1948). A nonparametric test of independence. *Annals of Mathematical Statistics* 26, 189-211.
- [18] Hong, Y. (1998). Testing for pairwise serial independence via the empirical distribution function. *Journal Royal Statistical Society* 60, 429-453.
- [19] Kauppi, H. and P. Saikkonen (2008). Predicting U.S. recessions with dynamic binary response models. *Review of Economics and Statistics* 90, 777-791.
- [20] Kheifets, I.L. (2011). Specification tests for nonlinear time series. Mimeo.
- [21] Khmaladze, E.V. (1981). Martingale approach in the theory of goodness-of-tests. *Theory of Probability and its Applications* 26, 240-257.

- [22] Koul, H.L. and W. Stute (1999). Nonparametric model checks for time series. *Annals of Statistics* 27, 204-236.
- [23] Mora, J. and A.I. Moro-Egido (2007). On specification testing of ordered discrete choice models. *Journal of Econometrics* 143, 191-205.
- [24] Neslehova, J. (2006). *Dependence of Non Continuous Random Variables*. Springer-Verlag.
- [25] Phillips, P.C.B. and J.Y. Park (2000). Nonstationary Binary Choice. *Econometrica* 68, 1249-1280.
- [26] Politis, D., J. Romano and M. Wolf (1999). *Subsampling*. New York: Springer-Verlag.
- [27] Rosenblatt, M. (1975) A quadratic measure of deviation of two-dimensional density estimates and a test of independence. *Annals of Statistics* 3, 1-14.
- [28] Rydberg, T.N. and N. Shephard (2003). Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics* 1, 2-25
- [29] Shao, J. and T. Dongsheng (1995). *The Jackknife and bootstrap*. New York: Springer-Verlag.
- [30] Skaug, H.J. and D. Tjøstheim (1993). Nonparametric test of serial independence based on the empirical distribution function. *Biometrika* 80, 591-602.
- [31] Startz, R. (2008). Binomial Autoregressive Moving Average Models with an Application to U.S. Recessions. *Journal of Business and Economic Statistics* 26, 1-8.
- [32] Wooldridge, J.M. (1990). An encompassing approach to conditional mean tests with applications to testing nonnested hypotheses. *Journal of Econometrics* 45, 331-350.